Avaliação de cursos de graduação

Claudio de Moura Castro

Simon Schwartzman

O sistema de avaliação de cursos do MEC começou com o Provão e se transformou no CONAES, um ente algo indefinido. Não cumpre aqui mapear os meandros da sua evolução e nem os seus bastidores. Examinamos o sistema atual, com suas evoluções mais recentes. O último capítulo dessa evolução é o CPC, um índice composto do ENADE (o sucessor do Provão), do IDD (uma medida de valor adicionado) e das variáveis de processo (corpo docente, regime de trabalho etc.).

Antes em entrar em matéria, proclamamos em alto e bom tom a nossa apreciação ao MEC e INEP por dar seguimento à uma política agressiva de avaliação do ensino e ampla divulgação dos resultados. Nossas críticas são muito específicas e se referem a aspectos que não comprometem os méritos do que está sendo feito nessa área.

Na primeira parte do trabalho, apresentamos uma crítica à nova fórmula. Em seguida, oferecemos uma proposta para retocar esse sistema.

Uma crítica ao CPC

Temos as seguintes objeções ao CPC:

i) *O uso inapropriado de um índice composto*

Nem sempre é boa idéia somar alhos com bugalhos. Índices compostos podem ser apropriados em certos casos e impróprios em outros.

O IDEB é um índice composto. E podemos defender o seu uso. O IDH também e nada temos contra ele.

Por que então impugnar o novo indicador do MEC? É simples, uma mensuração é uma resposta a uma necessidade de uso. O IDH permite comparar países ou cidades. Agrupa renda, saúde e educação, fatores reconhecidos que medem diferentes aspectos da qualidade de vida. Em conjunto, dão uma idéia melhor do conceito complexo que se está tentando medir.

Mas ao mesclar o que saem sabendo os graduados (ENADE) com a contribuição líquida do curso (IDD), estamos justamente obliterando diferenças que interessa conhecer. Tomemos o exemplo de dois cursos de farmácia incluídos no CPC. Um empregador gostaria de saber que o primeiro curso de farmácia obteve 5 (quanto o graduado sabe) e 2 (quanto o curso adicionou ao conhecimento do aluno). Ou seja, excelente qualidade final e lastimável adição ao conhecimento dos alunos. Contratar alguém desse curso pode ser uma boa idéia, pois sabe mais. Contudo, para um aluno modesto que deseja um curso universitário que o faça crescer, esse é uma péssima escolha.

Em contraste, há outro curso que obteve 2 e 5. Para esse mesmo aluno modesto que deva escolher onde fazer seu vestibular, o segundo curso dá a ele maior perspectiva de crescimento pessoal.

Se o objetivo do MEC é apenas decidir inicialmente que cursos deixar correr soltos e quais "internar no CTI", bastaria o resultado na prova aos graduandos (ENADE). Se o escore é muito baixo, sinal de alarme: é preciso ver o que está errado. Ainda no caso dos dois cursos citados, o segundo tem méritos, pois alavanca as carreiras pessoais dos alunos, apesar de que entram muito mal preparados. Já o ensino do primeiro curso é péssimo, pois recebe alunos bons e pouco oferece a eles. O que fazer com cada um? Essas são decisões críticas para o MEC: Punir um curso fraco na sala de aula, mas que recebe bons alunos? Punir um curso que faz um bom trabalho, mas recebe alunos fracos? O índice que junta os dois indicadores oblitera lacunas muito diferentes.

ii) *Tal como formulado, o índice introduz um viés a favor do ensino público*

Em uma pesquisa em co-autoria de José Francisco Soares, com um dos autores da presente nota, foi possível medir com considerável precisão o valor adicionado. As equações nos permitiram também concluir que da ordem de 80% da variância explicada se deveu à pontuação dos alunos ao entrarem no superior. Ou seja, sobram apenas 20% que são explicados pela excelência ou mediocridade do ensino oferecido.

A vasta maioria dos cursos superiores públicos recebe os melhores alunos. Esse fato, em si mesmo, garante um resultado satisfatório, quaisquer que sejam os méritos do ensino que oferece. Isso é um fato, a prova mede o resultado final e é tudo que pode fazer. Contudo, somar à prova dos graduandos o resultado da prova aplicada aos calouros infla artificialmente o resultado do ENADE para os cursos que recebem os melhores alunos. Se o curso ensina mal, mas, recebe alunos bons, a medida introduz um viés que esconde a fragilidade da sua sala de aula, superestimando, *ipso facto*, a qualidade do curso. Por que o MEC estaria usando uma medida tão bizarra?

iii) *Ao tomar medidas de resultado e juntar a elas medidas de processo, o índice mescla meios com fins*

Imaginemos engenheiros que precisam avaliar o desempenho de um protótipo de avião. Depois do avião voar, basta medir o seu desempenho no ar. Mas enquanto ainda não voou, como não há medidas de resultados, são obrigados a todos os malabarismos teóricos para prever como a aeronave se comportará após a decolagem.

Assim são as avaliações americanas. Se lá houvesse um Provão ou ENADE, poderiam jogar fora as dezenas de indicadores de processo que são obrigados a usar. Isso seria um alívio, pois é sabido que são muito imperfeitos como preditores de desempenho. Mas como são o que existe, os americanos têm que usá-los. Só o Brasil tem uma prova desse tipo, aplicada em graduados. Portanto, podemos e devemos dispensar as medidas de processo.

Se aplicarmos uma prova para medir o que aprenderam os graduados, como chegaram lá se torna irrelevante. Isso permite evitar a introdução de vieses. Por exemplo, somente cursos públicos têm recursos para manter todos ou quase todos os professores em tempo integral. Independentemente do que saem sabendo os alunos, conceder aos públicos um bônus de pontos no escore final só se justificaria se não tivéssemos uma prova medindo o resultado final.

Igualmente, um curso privado que contrata profissionais atuantes no mercado, estará oferecendo um ensino melhor nas disciplinas aplicadas – comparado com as públicas. Não obstante, será penalizado na nota final.

Já foi apresentado o argumento de que a prova ainda é imperfeita. Sendo assim, deveria ser reforçada com outros indicadores. Mas tal como vemos a questão, se esse é o caso, cumpre melhorar a prova, e não introduzir indicadores de processo, de validade ainda mais duvidosa.

Naturalmente, se a nota dos graduados é baixa demais, nesse caso, vamos usar as variáveis de processo, para identificar onde pode estar o problema. Os meios ou os processos são variáveis de diagnóstico de disfunções. São eminentemente úteis para isso.

Usando uma metáfora de restaurantes franceses, o visitador do Michelin não se ocupa de saber a marca do fogão e nem os contratos de trabalho dos cozinheiros. Apenas conta a qualidade da comida e do serviço que observaram. Somente se os clientes sofreram uma epidemia de salmonela, é que a saúde pública vai verificar se os pratos e a cozinha foram rigorosamente esterilizados. Mas isso é outro assunto. Não se mescla com a avaliação dos resultados finais.

iv) *Não conhecemos bem a natureza dos indicadores usados*

Não amadurecemos ainda uma boa interpretação dos resultados do valor adicionado (IDD). É um conceito novo e de difícil interpretação. Não estamos diante de um construto com interpretações intuitivas e transparentes. Daí ser prematuro fundi-lo com o ENADE ou com outros indicadores.

Caberia, nesse momento, explorar o IDD e mostrar como se combina com o ENADE, tomando como exemplo a observação dos cursos incluídos nas avaliações já realizadas. Pergunte-se a qualquer jornalista da área se o IDD dos cursos muito bons tende a ser maior ou menor do que o dos cursos fraquíssimos. Provavelmente, não sabem, embora essa diferença seja importantíssima para a política pública. Responde a uma pergunta crucial: os cursos de desempenho fraco dos graduandos estão oferecendo pouco aos seus alunos? Ou podem estar oferecendo muito a alunos fracos? Tal como o conceito de “empate técnico” nas pesquisas de intenção de voto, o IDD leva tempo para ser digerido pela opinião pública.

O mesmo desconhecimento existe para os indicadores de insumos ou de processo. Muitos livros na biblioteca fazem os alunos aprender mais? Professores de tempo integral têm um impacto positivo no aprendizado? E nas áreas profissionais? Doutores ensinam melhor? Como se comparam com mestres? Para que perfil de alunos? Mesmo as análises multivariadas são muito enganosas, por confundir causa, efeito e multicolinearidade.

Daí a minha crítica ao MEC, por divulgar um conceito excessivamente complexo, abstrato e cheio de cacoetes. Na prática, o que vimos na imprensa foi o previsível: Oba! Mais uma olimpíada do MEC. Quem são os medalhistas? Quem são os fracassados? Ao mesmo tempo, não houve esforços de interpretar as nuances dos resultados.

Mudanças sugeridas

Em linhas gerais, a proposta que defendemos tem como eixo as proposições que seguem:

i) Abandonar os indicadores compostos. Focalizar as análises e a disseminação dos dados em dois indicadores separados, o ENADE e o IDD.

ii) As variáveis de processo somente entram nas análises subseqüentes, quando se busca um diagnóstico de problemas identificados

Examinamos adiante detalhes das nossas sugestões.

Enade

O ENADE, sucessor do provão, constitui-se no elemento âncora de todo o processo de avaliação. Afinal de contas, nada mais central do que medir o conhecimento adquirido pelos alunos ao sair do curso. Por essa razão, justifica-se tudo que se possa fazer para o seu aperfeiçoamento.

Propomos as seguintes modificações:

i) Mudar a periodicidade das provas (tenho dúvidas aqui. Há uma questao de custos, e eu tenderia a pensar que os processos de mudança tendem a ser lentos. Tres anos me parece de bom tamanho.

Um exame anual cumpre muito melhor sua função de sinalizar para o grande público a qualidade do graduado egresso de cada curso. O intervalo de três anos deixa no limbo, por muito tempo, cursos que melhoram após receber um resultado negativo na prova. E prejudica aos alunos que poderiam querer evitar cursos atingidos por uma crise brusca.

Se há realmente obstáculos do lado orçamentário, é possível reduzir os custos. Uma alternativa é avaliar todos os anos os cursos no terço inferior da distribuição, a cada dois anos o terço do meio e a cada três anos o superior. Como estamos propondo também eliminar a prova no primeiro ano, virá por aí mais uma redução de custos.

ii) Incluir o universo dos alunos

Abandonar a prática da amostragem e passar a incluir todos os graduados das carreiras avaliadas. O custo de aplicar uma prova à metade ou a todos os alunos de uma turma é praticamente o mesmo. Além de reduzir o risco de fraude, essa medida irá gerar resultados individuais. Todos os alunos saberão a nota obtida e podem usar isso em seu benefício. Além disto, eliminará as dúvidas que persistem sobre a representatividade das amostras.

iii) Fazer a nota constar do diploma do aluno

Com isso, evitam-se os boicotes por parte dos alunos. Além disso, cria-se um indicador individual de desempenho, que pode ser um passaporte no mercado de trabalho, ou para a entrada na pós-graduação. (O problema aí é a qualidade da prova. Se ela for excelente, seria o equivalente a um exame de ordem, ou certificaçao. Mas nao se pode fazer isto em uma prova mal feita e superficial – aí eu concordo com o comenário do Verhine. A outra coisa é que as provas nao sao comparáveis de um ano para outro, e o significado de ter uma pontuaçao determinada pode ser muito diferente de ano a ano.

Há também outro uso potencialmente muito útil. Sabemos que a variância de desempenho dentro de cada curso é muito grande. Um bom aluno de um curso fraco pode ser amplamente superior a um aluno fraco em um curso de alta reputação. Se as notas são oferecidas aos alunos, torna-se possível para os melhores alunos das instituições mais fracas demonstrar a todos (em particular, empregadores e cursos de pós-graduação) que seu desempenho é superior.

iv) Eliminar a aplicação da prova no primeiro ano

Como dito anteriormente, a prova no primeiro ano serve para calcular o valor adicionado, além de ser somada ao escore no último ano. Não vemos razões ponderáveis para somar à prova do último ano o que sabiam os alunos no primeiro. E como explicaremos com mais vagar na próxima seção, o ENEM dos entrantes em cada curso resolve o problema de calcular o valor adicionado. Portanto, esta prova pode ser eliminada, sem perda para a avaliação e com consideráveis reduções de custos

v) Introduzir matrizes de competências para a formulação das provas

Tanto o ENADE quanto o Provão jamais utilizaram uma prova formulada de acordo com as melhores práticas na área. Não conhecemos erros óbvios nas provas. Contudo, são formuladas sem uma matriz de competências que permita cobrir de forma sistemática os tópicos mais centrais do perfil profissional correspondente.

Por outro lado, é preciso verificar até que ponto as provas nas áreas profissionais refletem as melhores práticas da profissão. Há um problema crônico, descrito há muitos anos. É o que Ronald Dore chamou de “academic drift”. Trata-se da tendências interna das instituições para esvaziar as competências profissionais e substituí-las pelo peso das áreas acadêmicas. É sempre necessário impedir que se alastre tal tendência. Isso significa, pelo menos, que na formulação das matrizes e na redação das provas haja a participação de profissionais do mercado.

vi) Eliminar a prova de “cultura geral”

A prova de cultura geral pode parecer, à primeira vista, uma boa idéia. Mas se a formulação de uma matriz de competências para uma prova profissional já oferece certos desafios, fazer o mesmo com “cultura geral” é muito mais difícil. O tema é tão amplo que se torna quase impossível definir o que deve saber o aluno nessa área. Literatura? Filosofia? História da Arte? O grande cenário da política mundial?

Em princípio, não sabemos bem o que consegue medir uma prova destas. O melhor que se pode fazer é eliminar tal prova, pois nem mede alguma coisa que tenha um sentido claro e nem orienta o esforço do aluno em uma direção produtiva.

vii) Considerar a possibilidade de aumentar o número de questões

Visando melhorar a cobertura das matérias consideradas e para aumentar a fidedignidade dos resultados pode fazer sentido aumentar o número de questões na prova.

viii) Voltar ao procedimento de dar notas na curva normal.

Como não há parâmetros de performance para a prova, não é possível saber se os resultados revelam alunos que sabem pouco ou uma prova difícil demais. Ademais, o uso de notas em valor absoluto convida a comparações espúrias entre diferentes áreas. O provão evitava tais comparações pelo uso da curva normal como critério de notas. Era "A" quem estava no segmento acima do percentil 88.

Mesmo que a imprensa custasse a entender esse valor puramente relativo, pelo menos, não se podia dizer que "matemática se saiu pior que serviço social". O grande problema aqui é a inexistência de uma definição oficial e confiável do que seriam os conhecimentos esperados dos graduados de cada curso superior. Sem isso, não se pode nem sequer pensar em criar níveis mínimos aceitáveis.

O erro do sistema presente é dar a impressão de que tais critérios existem. Um curso com nota baixa é considerado como ruim, como se houvesse uma métrica oficial. E se for argüido que tal métrica existe, o remendo fica pior do que o soneto, pois implica que um bando de formuladores de perguntas reunidos *ad hoc* recebeu um mandado oficial de estabelecer tais níveis.

Voltando o critério para curva normal, recuperamos a prática muito mais sadia de dizer que o curso A é melhor do que o B, nada mais sendo possível afirmar.

(pelo que consegui entender, as notas do ENADE tambem sao normalizadas, só que a maneira pela qual eles tratam os extremos faz com que as médias fiquem deslocadas.

Valor adicionado (IDD)

O valor adicionado pelo curso traz uma nova variável muito relevante. Responde à pergunta: o curso é bom porque recebe alunos bem preparados ou porque oferece muito aos seus alunos, independente do quanto sabem ao entrar?

A sua introdução foi um avanço considerável. Não obstante, o seu cálculo foi proposto em um momento em que o ENEM já existia e já era aplicado em um universo suficientemente grande para obter significância estatística, na vasta maioria dos casos.

Reaplicar a mesma prova de saída ao fim do curso foi uma decisão inexplicável. Na época, o ENEM estava em desfavor no INEP e corriam boatos de que seria extinto. Seria essa a razão?

Agora que o ENEM se aplica a um universo enorme, maiores razões há para usá-lo como critério de desempenho de entrada no superior.

Há várias maneiras de estimar o valor adicionado. Na mais sofisticada, a cada medida de rendimento individual na entrada corresponde uma medida na saída, para a mesma pessoa.

Mas isso não é necessário, para que se obtenham resultados bastante confiáveis. Podemos considerar a média da prova de entrada para a turma e a média na prova de saída. Desta forma, podemos calcular o valor adicionado do curso como um todo. No final das contas, é o resultado que nos interessa.

No caso do ENEM, podemos começar com valores médios para a turma e progressivamente criar mecanismos para pescar o ENEM original de cada formando – ou grupo de formandos – que faz a prova do ENADE.

Isto nao está claro para mim. Vai se comparar a média do ENEM na entrada com a méda da prova específica ao final? Vai se fazer outra prova do ENEM na saída? Quas sao as suposiçoes de conteúdo em relaçao a isto?

Variáveis de processo

Como sugerido na primeira parte da presente nota, uma avaliação tem muitos objetivos, cada um sugerindo uma fórmula diferente de medição. Uma óbvia utilidade das provas é informar alunos, cursos e empregadores acerca do nível de conhecimento dos alunos de cada instituição. Para isso, basta o ENADE, complementado pelo valor adicionado.

Não faz sentido dar mais pontos a um curso que tem essa ou aquela característica de funcionamento. De fato, se tal característica for eficaz, o seu efeito de melhorar as notas dos estudantes já está incluído na pontuação obtida.

Se o objetivo é interno do MEC, para identificar cursos que merecem atenção especial, tampouco é necessário entrar com variáveis de processo. Uma vez identificados os cursos abaixo de certo limiar de qualidade, aí então, entram em cena as variáveis de processo.

Por que o curso é fraco? Seus professores não têm o perfil acadêmico requerido? Que outras falhas se poderão encontrar?

Na verdade, estamos longe de conhecer o real impacto dessas variáveis de processo. Tome-se a relação aluno-professor. Trata-se de um tema exaustivamente pesquisado. E o que ficamos sabendo? Apenas que raramente se encontra alguma associação entre rendimento do aluno e essa variável.

Alunos de doutores aprendem mais? Nos Community Colleges dos Estados Unidos há uma firme convicção de que aprendem menos, pois há regras proibindo a sua contratação. Temos evidência de que essas instituições estão redondamente enganadas?

Mesmo quando se fazem análises de regressão com variáveis de processo, é enorme o risco de multi-colinearidade, passando por causação.

Enfim, dado o baixo nível de conhecimento que temos sobre o impacto de variáveis de processo, não parece ser o caso de usá-las em avaliações que são entregues ao grande público e que têm impacto na opinião pública.